AD-A103 884 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
THE ASYMPTOTIC DISTRIBUTION OF THE PRODUCT ESTIMATOR.(U)
AUG 81 C WU, D CHANG DAAG29-80-C-0041
UNCLASSIFIED MRC-TSR-2253 NL

END
DATE
FILMED
0 81
DTIC

AD A103884

MRC Technical Summary Report # 2253

THE ASYMPTOTIC DISTRIBUTION OF THE PRODUCT ESTIMATOR

C.-F. Wu and D.-S. Chang

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

See 1473

August 1981

(Received June 19, 1981)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

and

National Science Foundation
Washington, D. C. 20332

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

THE ASYMPTOTIC DISTRIBUTION OF THE PRODUCT ESTIMATOR

C.-F. Wu[1,2] and D.-S. Chang[*,1,3]

Technical Summary Report #2253

August 1981

ABSTRACT

The exact formulas for the bias and variance of the product estimator under simple random sampling are given. Its asymptotic normality is rigorously established under weak and interpretable regularity conditions on the finite populations. No superpopulation model is assumed. Hájek's projection method is the key tool in our proofs.

AMS (MOS) Subject Classification: 62D05, 62F12

Key Words: Product Estimator, Asymptotic Normality, Bias, Variance, Projection, Lindeberg-Hájek's Central Limit Theorem

Work Unit Number 4 (Statistics and Probability)

* National Tsing Hua University, Hsinchu, Taiwan.

1 Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.
2 Supported by the National Science Foundation under Grant No. MCS-7901846.
3 Supported by the National Science Council, Republic of China.

SIGNIFICANCE AND EXPLANATION

Suppose we want to estimate the average acres of corn per farm (denoted Y) in the state of Wisconsin based on a simple random sample of X values and available information on the size of each farm (denoted X). In this particular instance, X and Y are positively correlated. It is common to use the ratio estimator defined as (y-sample mean) • (x-population mean)/(x-sample mean) as the estimator of y-population mean. The ratio estimator is very simple to compute and is efficient for a general class of populations. In case X and Y are negatively correlated, the product estimator defined as (y-sample mean) • (x-sample mean)/(x-population mean) is more efficient for estimating the y-population mean than the ratio estimator. Its computation is as simple as the ratio estimator. Sometimes we want to find a confidence interval for the y-population mean. Since no exact finite-sample distribution theory is available, we study the asymptotic normality of the product estimator. Our regularity conditions are directly on the finite populations. No artificial superpopulation model is assumed. The conditions are weak and interpretable in practical terms so that the practitioners may find them useful.

THE ASYMPTOTIC DISTRIBUTION OF THE PRODUCT ESTIMATOR

C.-F. Wu[1,2] and D.-S. Chang[*,1,3]

1. Introduction

In sample surveys the ratio estimator $\hat{Y}_R = \bar{y}\bar{X}/\bar{x}$ using a suitable auxiliary variate X is one of the most commonly adopted estimates of the population mean of character Y under study. In case both variates Y and X are positive (or one positive and the other negative) and the corretlation coefficient ρ between Y and X has high negative value (or high positive value), the product estimator $\hat{Y}_p = \bar{y}\bar{x}/\bar{X}$ is preferred. Throughout the paper we assume the sample is obtained from simple random sampling (s.r.s). It is known (Murthy, 1964) that the product estimator has smaller mean square error than the unbiased estimator $\bar{y}$, if ρ is smaller than minus one-half of the ratio of the coefficients of variation of X and Y. Though the asymptotic behavior of ratio and regression estimators was discussed in Scott and Wu (1981), the corresponding results for the product estimator were not satisfactorily given in the literature. Srivastava's result (1966) is not rigorous, is for infinite population, and regularity conditions were not spelled out. Had the conditions been given there, they would be unnecessarily strong because his approach is not powerful enough. The technique of proof for the ratio estimator in Scott and Wu (1981) does not work for the product estimator. The key technical problem is that both $\bar{y}$ and $\bar{x}$ involve the same random sample. As explained in more detail in section 3, standard probability results can not be directly applied to this problem. The key

* National Tsing Hua University, Hsinchu, Taiwan.

1 Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.
2 Supported by the National Science Foundation under Grant No. MCS-7901846.
3 Supported by the National Science Council, Republic of China.

technical tool employed in this paper is Hájek's projection method which has proved successful for many other problems in statistics. We project the product variable $\bar{y}\,\bar{x}$ onto a linear space of random variables, thus obtaining a variable of simpler form. We then prove the asymptotic equivalence of the two variables. The asymptotic normality of the latter variable was proved by using the Lindeberg-Hájek Central Limit Theorem. The required regularity conditions are weak and easily interpretable.

In the next section we derive exact formulas for the bias and variance of the product estimator. The asymptotic normality of the product estimator is proved in section 3, using the projection method outlined in the previous paragraph. The asymptotic consistency of a variance estimator is established in section 4. Finally the two results are combined to give the main Theorem 3, which is followed by some remarks on regularity conditions.

2. Bias and Variance of the Product Estimator

Before discussing the asymptotic results of the product estimator, we first derive the exact formulas of the bias and variance of the product estimator and use the latter to obtain an approximate variance in the large sample case. The technique is the same as in Goodman (1960) or Sukhatme and Sukhatme (1970, p. 190-192).

In population $U = \{1,2,\ldots,N\}$, let $f = n/N$ be the sampling fraction, $\bar{y}$ and $\bar{x}$ be the sample means of simple random sample of size n of the variates Y and X respectively, and $\bar{Y}$ and $\bar{X}$ be the corresponding population means. Denote

$$S_{\alpha\beta} = (N-1)^{-1} \sum_{i=1}^{N} (Y_i-\bar{Y})^{\alpha}(X_i-\bar{X})^{\beta}$$

and (1)

$$C_{\alpha\beta} = S_{\alpha\beta}/\bar{Y}^{\alpha}\bar{X}^{\beta} \ , \quad \alpha,\beta = 0,1,2 \ .$$

Then, we have the following

Lemma 1.

(i) $E(\bar{y}\bar{x}) = \bar{Y}\bar{X}(1 + \frac{1-f}{n} C_{11})$

and

(ii)
$$\mathrm{Var}(\bar{y}\bar{x}) = (\bar{Y}\bar{X})^2(\frac{1-f}{n})\Big\{C_{20}+C_{02}+2C_{11} + \frac{1}{n}\Big[2(\frac{N-2n}{N-2}) \cdot (C_{12}+C_{21})$$

$$+ \frac{(n-1)(N-1)(N-n-1)}{n(N-2)(N-3)} (C_{02}C_{20}+C_{11}^2) \qquad (2)$$

$$- \frac{N-n}{N} C_{11}^2\Big] + \frac{1}{n^2} \cdot \frac{N^2+N-6nN+6n^2}{(N-2)(N-3)} C_{22}\Big\} .$$

Proof.

Let

$$\delta\bar{x} = (\bar{x}-\bar{X})/\bar{X} \ , \qquad \delta\bar{y} = (\bar{y}-\bar{Y})/\bar{Y} \ . \qquad (3)$$

Then, using Theorem 2.3 of Cochran (1977), we have

$$E(\bar{y}\bar{x}) = \bar{Y}\bar{X}[1+E(\delta\bar{y}\delta\bar{x})] = \bar{Y}\bar{X}(1 + \frac{1-f}{n} C_{11}) .$$

Using results in Sukhatme and Sukhatme (1970, p. 190-192) and after some simplication, we have

$$\mathrm{Var}(\bar{y}\bar{x}) = E(\bar{y}\bar{x})^2 - [E(\bar{y}\bar{x})]^2$$

$$= (\bar{Y}\bar{X})^2 E[(\delta\bar{y})^2(\delta\bar{x})^2+2(\delta\bar{y})^2\delta\bar{x}+2\delta\bar{y}(\delta\bar{x})^2+(\delta\bar{y})^2+(\delta\bar{x})^2$$

$$+4\delta\bar{y}\delta\bar{x}+2\delta\bar{y}+2\delta\bar{x}+1] - [E(\bar{y}\bar{x})]^2$$

$$= (\bar{Y}\bar{X})^2(\frac{1-f}{n}) \Big\{C_{20}+C_{02}+2C_{11} + \frac{1}{n}\Big[2(\frac{N-2n}{N-2})(C_{21}+C_{12})$$

$$+ \frac{(n-1)(N-1)(N-n-1)}{n(N-2)(N-3)} (C_{20}C_{02}+2C_{11}^2) - \frac{N-n}{N} C_{11}^2\Big]$$

$$+ \frac{1}{n^2} \cdot \frac{N^2+N-6nN+6n^2}{(N-2)(N-3)} C_{22} \Big\}.$$

Applying the results in Lemma 1, we have the following formulas for the product estimator $\hat{\bar{Y}}_p = \bar{y}\bar{x}/\bar{X}$.

(i) $\quad E(\hat{\bar{Y}}_p) - \bar{Y} = \frac{1-f}{n} C_{11}\bar{Y}$

and

(ii) $$\mathrm{Var}(\hat{\bar{Y}}_p) = \bar{Y}^2(\frac{1-f}{n}) \Big\{C_{20}+C_{02}+C_{11}$$

$$+ \frac{1}{n}\Big[2(\frac{N-2n}{N-2})(C_{21}+C_{12}) + \frac{(n-1)(N-1)(N-n-1)}{n(N-2)(N-3)}$$

$$\times (C_{20}C_{02}+2C_{11}^2) - \frac{N-n}{N} C_{11}^2\Big] \qquad (4)$$

$$+ \frac{1}{n^2} \frac{N^2+N-6nN+6n^2}{(N-2)(N-3)} C_{22} \Big\}$$

The formula $\mathrm{Var}(\overline{\overline{yx}})$ in Goodman (1960, (21), p. 712) differs from ours in that Goodman's treatment of product is for sampling from infinite population. We should point out that the coefficient of E_{11}^2 in his formula (21) should be n^{-2} instead of n^{-3}, a minor error which disappears as $n \to \infty$.

3 The Asymptotic Distribution of the Product Estimator

Following the usual formulation of the central limit theorem, we embed our finite population in a sequence of populations, $\{U_v\}$, indexed by v where n_v and N_v both increase without bound as $v \to \infty$. And let $\{s_v\}$ be a sequence of simple random samples with sample size n_v and sample means $\bar{y}_v$, $\bar{x}_v$ of Y and X respectively in s_v. Similarly, let $\bar{Y}_v$ and $\bar{X}_v$ be the population means of Y and X, S^2_{vy} the population variance of Y in U_v, and $f_v = n_v/N_v$ the sampling fraction. All our results are based on the following two standard theorems in finite populations (see Scott and Wu, 1981). We use $\xrightarrow{L}$ and $\xrightarrow{P}$ to denote convergence in distribution and in probability respectively.

<u>Theorem A.</u>

Suppose $n_v \to \infty$ and $N_v - n_v \to \infty$ as $v \to \infty$ Then, under simple random sampling,

$$\frac{\sqrt{n_v}\,(\bar{y}_v - \bar{Y}_v)}{\sqrt{1-f_v}\; S_{vy}} \xrightarrow{L} N(0,1) \quad \text{as} \quad v \to \infty$$

if and only if $\{y_{vj}\}_{v,j}$ satisfies the Lindeberg-Hájek condition

$$\lim_{v\to\infty} \sum_{T_v(\varepsilon)} \frac{(y_{vj} - \bar{Y}_v)^2}{(N_v-1)S^2_{vy}} = 0 \quad \text{for any} \quad \varepsilon > 0 , \qquad \text{(A)}$$

where $T_v(\varepsilon)$ is the set of units in U_v for which

$$\frac{|y_{vj} - \bar{Y}_v|}{\sqrt{1-f_v}\; S_{vy}} > \varepsilon\sqrt{n_v} .$$

Theorem B.

Suppose $\{y_{\nu y}\}$ satisfies the condition

$$(1-f_\nu)S^2_{\nu y}/n_\nu \to 0 \quad \text{as} \quad \nu \to \infty \tag{B}$$

Then, under simple random sampling,

(i) $E\,|\bar{y}_\nu - \bar{Y}_\nu| \to 0$

and

(ii) $\bar{y}_\nu - \bar{Y}_\nu \xrightarrow{P} 0$ as $\nu \to \infty$.

The main technical problem stems from the fact that both $\bar{y}$ and $\bar{x}$ involve the same set of indicator variables I_i, further complicated by the dependency of I_i among $i = 1,2,\ldots,N$. As far as we know, no existing results in probability theory can be directly applied to our problem. We project the variable $\xi_\nu = n^2_\nu \bar{y}_\nu \bar{x}_\nu$ onto the space of linear combinations of $\{I_i\}^N_{i=1}$ and establish the asymptotic equivalence of ξ_ν and its projection. Since the projection variable is a weighted sum of I_i, its asymptotic normality can be established relatively easily.

For simplicity we drop the subscript ν in all the derivations and proofs.

Let I_i be the indicator of unit i in simple random sample s without replacement of size n. We can write the centered version of the product variable $\xi = n^2\bar{x}\bar{y}$ as

$$\eta = \xi - E(\xi) = \sum_{i=1}^{N}\sum_{j=1}^{N} X_iY_j[I_iI_j - E(I_iI_j)] \ . \tag{5}$$

We project η onto the space Z of linear combinations of $\{I_i\}^N_{i=1}$, i.e.,

$$Z = \{z \mid z = \sum_{i=1}^{N} k_iI_i,\ k_i \text{ is real-valued constant for } i = 1,\ldots,N\}$$

Let the projection variable be W. Then we have (Lehmann, 1975, p. 362),

$$W = \sum_{i=1}^{N} E(\eta|I_i) \; . \tag{6}$$

By taking conditional expectation, we have the following properties, where $f = n/N$ and $f_1 = (n-1)/(N-1)$,

$$\text{(i)} \quad E(I_i) = f \quad , \qquad \text{(ii)} \quad E(I_i I_j) = ff_1$$

$$\text{(iii)} \quad E(I_j|I_i) = \frac{1}{N-1}(n-I_i), \qquad \text{(iv)} \quad E(I_k I_j|I_i) = f_1 \frac{1}{N-2}(n-2I_i) \tag{7}$$

for $i \neq j \neq k = 1,2,\ldots,N$.

From (7) we have (for details, see Chang, 1981)

$$E(\eta|I_i) = \left[f_1 \frac{N^2}{N-2} \bar{X}\bar{Y}(\delta x_i + \delta y_i) + \frac{N(N-2n)}{(N-1)(N-2)}(X_i Y_i - \frac{1}{N} \sum_k X_k Y_k) \right] (I_i - f) \; ,$$

for $i = 1,2,\ldots,N$ and from (6), we obtain

$$W = \sum_{i=1}^{N} a_i (I_i - f) = \sum_{i=1}^{N} (b_i + c_i)(I_i - f) \tag{8}$$

where $a_i = b_i + c_i$ and

$$b_i = f_1 \frac{N^2}{N-2} \bar{X}\bar{Y}(\delta x_i + \delta y_i) \; ,$$

$$c_i = \frac{N(N-2n)}{(N-1)(N-2)}\left(X_i Y_i - \frac{1}{N} \sum_{k=1}^{N} X_k Y_k\right) \; , \tag{9}$$

and

$$\delta x_i = \frac{X_i - \bar{X}}{\bar{X}} \quad , \quad \delta y_i = \frac{Y_i - \bar{Y}}{\bar{Y}} \quad , \quad i = 1,2,\ldots,N \; .$$

Next, by (2.26) and (2.27) in Cochran (1977), we have

$$\mathrm{Var}(W) = \sum_{i=1}^{N} (b_i + c_i)^2 \, \mathrm{Var}(I_i) + \sum_{i \neq j}^{N}\sum (b_i + c_i)(b_j + c_j)\mathrm{Cov}(I_i, I_j)$$

$$= (f - f^2) \left[\sum_{i=1}^{N} (b_i^2 + c_i^2 + 2b_i c_i) - \frac{1}{N-1} \sum_{i \neq j}^{N}\sum (b_i b_j + c_i c_j + 2b_i c_j) \right] .$$

From $\sum_{i=1}^{N} b_i = \sum_{i=1}^{N} c_i = 0$ and (9), we have

$$\sum_{i=1}^{N} b_i^2 = - \sum_{i\neq j}^{N}\sum b_i b_j = \left[f_1 \left(\frac{N^2}{N-2}\right)\bar{X}\bar{Y}\right]^2 (N-1)(C_{20}+C_{02}+2C_{11}) \ ,$$

$$\sum_{i=1}^{N} c_i^2 = - \sum_{i\neq j}^{N}\sum c_i c_j = \left[\frac{N(N-2n)}{(N-1)(N-2)}\bar{X}\bar{Y}\right]^2 (N-1) \tag{10}$$

$$\times \left[C_{22} + \frac{(N-1)(N-2)}{N^2} C_{11}^2 + 2(C_{21}+C_{12}+C_{20}+C_{02} + \frac{3N-2}{N} C_{11} + 1)\right]$$

and

$$\sum_{i=1}^{N} b_i c_i = - \sum_{i\neq j}^{N}\sum b_i c_j = f_1 \frac{N^3(N-2n)}{(N-2)^2} \bar{X}^2\bar{Y}^2 (C_{12}+C_{21}+C_{20}+C_{02}+2C_{11}) \ .$$

Derivation of the identities in (10) follows easily from definition (9) and formulas in Sukhatme and Sukhatme (1970, p. 190-192). Thus, finally we obtain

$$\begin{aligned} \mathrm{Var}(W) = n^3(1-f)\bar{X}^2\bar{Y}^2 \Bigg\{ & \left[\frac{(n-1)N^2}{n(N-1)(N-2)}\right]^2 (C_{20}+C_{02}+2C_{11}) \\ & + \frac{2}{n}\left[\frac{(n-1)(N-2n)N^3}{n(N-1)^2(N-2)^2} (C_{21}+C_{12}+C_{20}+C_{02}+2C_{11})\right] \\ & + \frac{1}{n^2}\left[\frac{N(N-2n)}{(N-1)(N-2)}\right]^2 \left[C_{22}+ \frac{(N-1)(N-2)}{N^2} C_{11}^2 \right. \\ & \left. + 2\left(C_{21}+C_{12}+C_{20}+C_{02}+ \frac{3N-2}{N} C_{11}+1\right)\right]\Bigg\} \ . \end{aligned} \tag{11}$$

Further, by (5) and (6), we can decompose

$$\eta = W + Q$$

and verify the identity

$$Var(\eta) = Var(W) + Var(Q). \quad (12)$$

Derivation of (12) is similar to that in Lehmann (1971, A. 168, p. 363). For details see Chang (1981, Lemma 4.2).

Next, we want to demonstrate the asymptotic equivalence between η and W as follows.

<u>Lemma 2</u>.

$Var(\eta_v)/Var(W_v) \to 1$ as $v \to \infty$ provided

(i) n_v and N_v both $\to \infty$ as $v \to \infty$,

(ii) $\bar{X}_v$, $\bar{Y}_v$ and $C_{v\alpha\beta}$, $\alpha,\beta=0,1,2$ are all uniformly bounded in v

and

(iii) $\{C_{v20}+C_{v02}+2C_{v11}\}_v$ is bounded away from zero uniformly in v, where $C_{v\alpha\beta}$ is defined in (1) for population U_v.

<u>Proof</u>.

Under conditions (i), (ii) and (iii), both $Var(\eta)$ and $Var(W)$ in (2) and (11) are dominated by $n^3\bar{Y}^{-2}\bar{X}^{-2}(C_{20}+C_{02}+2C_{11})$ as $v \to \infty$. This proves Lemma 2.

Similarly, under conditions (i), (ii) and (iii) of Lemma 2, both $Var[\sum_{i=1}^N b_i(I_i-f)]$ and $Var(W)$ are dominated by $n^3\bar{Y}^{-2}\bar{X}^{-2}(C_{20}+C_{02}+2C_{11})$ as $v \to \infty$. Therefore, we have

<u>Lemma 3</u>.

Under the same conditions as in Lemma 2,

$$Var(W_v)/Var[\sum_{i=1}^{N_v} b_{vi}(I_{vi}-f_v)] \to 1 \quad \text{as} \quad v \to \infty.$$

Now we let

$$W^* = \sum_{i=1}^{N} (\delta x_i + \delta y_i)(I_i - f) = \sum_{i \in s} (\delta x_i + \delta y_i)$$

and (13)

$$T(\varepsilon) = \left\{ i : \left| \frac{X_i}{\bar{X}} + \frac{Y_i}{\bar{Y}} - 2 \right| > \varepsilon\sqrt{Var(W^*)} \right\}, \quad \varepsilon > 0$$

Then, by applying the Lindeberg-Hájek Theorem A to W^* directly, we have

Lemma 4.

Let W_v^* and $T_v(\varepsilon)$ be defined in (13) for population U_v. Suppose that $n_v \to \infty$ and $N_v - n_v \to \infty$ as $v \to \infty$. Then

$$W_v^*/\sqrt{Var(W_v^*)} \xrightarrow{L} N(0,1) \quad \text{as} \quad v \to \infty$$

if and only if $\{X_{vj}/\bar{X}_v + Y_{vj}/\bar{Y}_v\}_{v,j}$ satisfies the Lindeberg-Hájek condition (A).

Combining Lemmas 2, 3 and 4, we establish the asymptotic distribution of the product estimator as follows

Theorem 1.

Under s.r.s,

$$\frac{\sqrt{n_v}\,(\hat{\bar{Y}}_{vp} - \bar{Y}_p)}{\sqrt{1-f_v}\;\bar{Y}_v\sqrt{C_{v20} + C_{v02} + 2C_{v11}}} \xrightarrow{L} N(0,1) \quad \text{as} \quad v \to \infty$$

provided

(i) $n_v \to \infty$ and $N_v - n_v \to \infty$ as $v \to \infty$,

(ii) $\{X_{vj}/\bar{X}_v + Y_{vj}/\bar{Y}_v\}_{v,j}$ satisfies the Lindeberg-Hájek condition (A),

(iii) $\bar{X}_v$, $\bar{Y}_v$ and $C_{v\alpha\beta}$, $\alpha,\beta = 0,1,2$ are all bounded uniformly in v

and

(iv) $\{C_{v20} + C_{v02} + 2C_{v11}\}_v$ is bounded away from zero uniformly in v.

Proof.

Decompose

$$\frac{\sqrt{n}\,(\hat{\bar{\bar{Y}}}_p-\bar{Y})}{\sqrt{1-f}\;\;\bar{Y}\sqrt{C_{20}+C_{02}+2C_{11}}} = \frac{\sqrt{n}}{\sqrt{1-f}\;\bar{Y}}\;\frac{(\hat{\bar{\bar{Y}}}_p-E\hat{\bar{\bar{Y}}}_p)}{\sqrt{C_{20}+C_{02}+2C_{11}}}$$

$$+\sqrt{\frac{1-f}{n}}\;\frac{C_{11}}{\sqrt{C_{20}+C_{02}+2C_{11}}}\;.$$

The second term converges to zero by condition (i), (iii), and (iv). The first term can be rewritten as

$$\frac{W+Q}{\sqrt{n^3(1-f)}\;\bar{\bar{X}}\bar{\bar{Y}}\;\sqrt{C_{20}+C_{02}+2C_{11}}} = \frac{W+Q}{\sqrt{Var(W)}}\;\frac{\sqrt{Var(W)}}{\sqrt{n^3(1-f)}\;\bar{\bar{X}}\bar{\bar{Y}}\;\sqrt{C_{20}+C_{02}+2C_{11}}}\;,$$

where, by conditions in (i), (iii), (iv), Lemma 2 and (12),

$$\sqrt{Var(W)}\;\sqrt{n^3(1-f)}\;\bar{\bar{X}}\bar{\bar{Y}}\;\sqrt{C_{20}+C_{02}+2C_{11}} \to 1$$

and

$$Q/\sqrt{Var(W)} \xrightarrow{P} 0 \quad \text{as} \quad v \to \infty.$$

Write $B = \sum_{i=1}^{N} b_i(I_i-f)$, $C = \sum_{i=1}^{N} c_i(I_i-f)$ and by (8), we have

$$\frac{W}{\sqrt{Var(W)}} = \frac{B}{\sqrt{Var(B)}}\cdot\sqrt{\frac{Var(B)}{Var(W)}} + \frac{C}{\sqrt{Var(W)}}\;.$$

Then, similarly, by conditions (i), (iii), (iv), Lemma 3 and (10),

$$C/\sqrt{Var(W)} \xrightarrow{P} 0$$

and

$$\sqrt{\mathrm{Var}(B)}/\sqrt{\mathrm{Var}(W)} \longrightarrow 1 \quad \text{as} \quad v \to \infty .$$

Since, by (9) and (13),

$$\frac{B}{\sqrt{\mathrm{Var}(B)}} = \frac{W^*}{\sqrt{\mathrm{Var}(W^*)}}$$

which converges in law to $N(0,1)$ by Lemma 4. The proof is completed by applying Slutsky's Theorem.

4. Variance Estimation

In practice, of course, $\bar{Y}^2(C_{20}+C_{02}+2C_{11})$ is unknown and is estimated by its sample analogue $\bar{y}^2(c_{20}+c_{02}+2c_{11})$. Let

$$v(\hat{\bar{Y}}_p) = \frac{1-f}{n}\bar{y}^2(c_{20}+c_{02}+2c_{11})$$

be an approximate variance estimate of $\mathrm{Var}(\hat{\bar{Y}}_p)$. Then, we prove the consistency of $v(\hat{\bar{Y}}_p)$ as follows.

Theorem 2.

Under s.r.s.,

$$v(\hat{\bar{Y}}_{vp})/(\frac{1-f_v}{n_v})\bar{Y}^2_v(C_{v20}+C_{v02}+2C_{v11}) \xrightarrow{P} 1 \quad \text{as} \quad v \to \infty$$

provided

(i) $n_v \to \infty$ and $N_v \to \infty$ as $v \to \infty$,

(ii) $\{(X_{vj}-\bar{X}_v)^2/S_{v02}\}_{v,j}$ and $\{(Y_{vj}-\bar{Y}_v)^2/S_{v20}\}_{v,j}$ both satisfy condition (B) of Theorem B.

(iii) $\bar{X}_v$, $\bar{Y}_v$ and $C_{v\alpha\beta}$, $\alpha,\beta=0,1,2$ are all bounded uniformly in v

and

(iv) $\{\rho_v^2\}$ and $\{C_{v20}+C_{v02}+2C_{v11}\}$ both are bounded away from zero uniformly in v, where ρ_v is the population correlation coefficient of X and Y in U_v.

Proof.

Write

$$v(\bar{Y}_p)/(\frac{1-f}{n})\bar{Y}^2(C_{20}+C_{02}+2C_{11})-1 = \frac{1}{(C_{20}+C_{02}+2C_{11})}\left\{\left[(\frac{\bar{y}}{\bar{Y}})^2c_{02}-C_{02}\right] + \left[(\frac{\bar{y}}{\bar{Y}})^2c_{20}-C_{20}\right] + 2\left[(\frac{\bar{y}}{\bar{Y}})^2c_{11}-C_{11}\right]\right\},$$

where,

$$\left(\frac{\bar{y}}{\bar{Y}}\right)^2 c_{02} - C_{02} = C_{02}\left[\frac{s_{02}}{S_{02}}\left(\frac{\bar{X}}{\bar{x}}\frac{\bar{y}}{\bar{Y}}\right)^2 - 1\right],$$

$$\left(\frac{\bar{y}}{\bar{Y}}\right)^2 c_{20} - C_{20} = C_{20}\left(\frac{s_{20}}{S_{20}} - 1\right)$$

and

$$\left(\frac{\bar{y}}{\bar{Y}}\right)^2 c_{11} - C_{11} = C_{11}\left(\frac{s_{11}}{S_{11}}\frac{\bar{X}}{\bar{x}}\frac{\bar{y}}{\bar{Y}} - 1\right) .$$

By condition (iii), the uniform boundedness of C_{v20} and C_{v02} imply that the variances of $\bar{x}/\bar{X}$ and $\bar{y}/\bar{Y}$ converge to 0. Together with the unbiasedness of $\bar{x}$ and $\bar{y}$,

$$\frac{\bar{X}}{\bar{x}} \xrightarrow{P} 1 \quad , \quad \frac{\bar{Y}}{\bar{y}} \xrightarrow{P} 1 \quad .$$

Further, by condition (ii) and Theorem B, we have

$$\frac{s_{02}}{S_{02}} \xrightarrow{P} 1 \ , \quad \frac{s_{20}}{S_{20}} \xrightarrow{P} 1 \ .$$

Finally, since

$$\frac{s_{11}}{S_{11}} - 1 = \frac{1}{\rho}\left(\frac{s_{11}}{\sqrt{S_{20}S_{02}}} - \rho\right) \quad ,$$

which converges to zero in probability by the uniform boundedness of $\{\rho^2\}$, condition (ii) and the proof of Theorem 4 of Scott and Wu (1981, p. 101.) Hence we complete the proof.

5. The Main Result and Some Remarks

By combining Theorems 1 and 2, we state the main result of the paper.

Theorem 3.

Under s.r.s,

$$\frac{\sqrt{n_v}(\hat{\bar{Y}}_{vp}-\bar{Y}_v)}{\sqrt{1-f_v}\ \bar{y}_v\ \sqrt{c_{v20}+c_{v02}+2c_{v11}}} \xrightarrow{L} N(0,1)$$

as $v \to \infty$ provided

(i) $n_v \to \infty$ and $N_v - n_v \to \infty$ as $v \to \infty$,

(ii) $\{X_{vj}/\bar{X}_v + Y_{vj}/\bar{Y}_v\}_{v,j}$ satisfies the Lindeberg-Hájek condition (A),

(iii) $\{(X_{vj}-\bar{X}_v)^2/S_{v02}\}_{v,j}$ and $\{(Y_{vj}-\bar{Y}_v)^2/S_{v20}\}_{v,j}$ both satisfy condition (B) of Theorem B.

(iv) $\bar{X}_v$, $\bar{Y}_v$ and $C_{v\alpha\beta},\ \alpha,\beta=0,1,2$ are all bounded uniformly in v

and

(v) $\{\rho_v^2\}$ and $\{C_{v20}+C_{v02}+2C_{v11}\}$ both are bounded away from zero uniformly in v.

Remarks.

Condition (ii) says that the contribution to the total sum of squares of the $(X_j/\bar{X} + Y_j/\bar{Y})$'s from gross outliers about their mean value should be relatively small. Condition (iii) on Y_j and X_j suggests, respectively, that the coefficient of variation of y and x should be reasonably small. Condition (iv) is weaker than the more typical assumption in the literature (e.g., David and Sukhatme, 1974; Krewski, 1978; Krewski and Rao, 1981): $\bar{X}_v, \bar{Y}_v, C_{v\alpha\beta} \to \bar{X}, \bar{Y}, C_{\alpha\beta}$ as $v \to \infty$. Condition (v) is assumed to avoid trivial cases and is satisfied in most practical situations.

References

Chang, D S. (1981). Estimation with Auxiliary Information in Survey Sampling: Some Large Sample Results. Ph.D. Thesis, Department of Statistics, University of Wisconsin-Madison, U.S.A.

Cochran, W. G. (1977). Sampling Techniques, 3rd edition, Wiley, New York.

David, I. P. and Sukhatme, B. V. (1974). On the bias and mean square error of the ratio estimator. Jour. Amer. Statist. Assoc. 69, 464-466.

Goodman, L. A. (1960). On the exact variance of products. Jour. Amer. Statist. Assoc. 55, 708-713.

Krewski, D. (1978). Jackknifing U-statistics in finite populations. Commun. Statist.-Theor. Meth. A7, 1-12.

Krewski, D. and Rao, J. N. K. (1981). Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods. Ann. Statist., to appear.

Lehmann, E. L. (1975). Nonparametrics: Statistical Methods Based on Ranks, Holden-Day, San Francisco.

Murthy, M. N. (1964). Product method of estimation. Sankhyā Ser. A 26, 69-74.

Scott, A. and Wu, C. F. (1981). On the asymptotic distribution of ratio and regression estimators. Jour. Amer. Statist. Assoc. 76, 98-102.

Srivastava, S. K. (1966). Product estimator. Jour.Ind. Statist. Assoc. 4, 29-37.

Sukhatme, P. V. and Sukhatme, B. V. (1970). Sampling Theory of Surveys With Applications, second edition. Iowa State University Press, Ames, Iowa, U.S.A.

C-FW, D-SC:lh

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: TSR# 2253
2. GOVT ACCESSION NO.: AD-A103884
3. RECIPIENT'S CATALOG NUMBER: (9) Technical
4. TITLE (and Subtitle): The Asymptotic Distribution of the Product Estimator
5. TYPE OF REPORT & PERIOD COVERED: Summary Report - no specific reporting period
6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): C.-F. Wu and D.-S. Chang
8. CONTRACT OR GRANT NUMBER(s): DAAG29-80-C-0041; NSF-MCS 7901846
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, Wisconsin 53706
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 4 (Statistics and Probability)
11. CONTROLLING OFFICE NAME AND ADDRESS: see item 18 below
12. REPORT DATE: August 1981
13. NUMBER OF PAGES: 17
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
15. SECURITY CLASS. (of this report): UNCLASSIFIED
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited.

(14) MRC-TSR-2253

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

and

National Science Foundation
Washington, D. C. 20332

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Product Estimator, Asymptotic Normality, Bias, Variance, Projection, Lindeberg-Hájek's Central Limit Theorem

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The exact formulas for the bias and variance of the product estimator under simple random sampling are given. Its asymptotic normality is rigorously established under weak and interpretable regularity conditions on the finite populations. No superpopulation model is assumed. Hájek's projection method is the key tool in our proofs.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 221200

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)